

Scaling of the energy gap in pattern-hydrogenated graphene

Roberto Grassi,^{1,2,*} Tony Low,^{1,†} and Mark Lundstrom¹

¹*School of Electrical & Computer Engineering, Purdue University, West Lafayette, Indiana 47906, USA*

²*ARCES and DEIS, University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy*

(Dated: November 22, 2011)

Recent experiments show that a substantial energy gap in graphene can be induced via patterned hydrogenation on an iridium substrate. Here, we show that the energy gap is roughly proportional to $\sqrt{N_H}/N_C$ when disorder is accounted for, where N_H and N_C denote concentration of hydrogen and carbon atoms, respectively. The dispersion relation, obtained through calculation of the momentum-energy resolved density of states, is shown to agree with previous angle-resolved photoemission spectroscopy results. Simulations of electronic transport in finite size samples also reveal a similar transport gap, up to 1 eV within experimentally achievable $\sqrt{N_H}/N_C$ value.

Introduction: Graphene's novel electronic and physical properties makes it an interesting material for various potential applications [1, 2]. For electronics, the absence of an energy gap limits its applicability. Currently, there are myriad known ways to opening an energy gap in graphene. For instance, the patterning of graphene nanoribbons induces a band gap due to the confinement of carriers along the transverse direction [3]. Although there are recent proposals in achieving controlled width and smooth edges [4], the large scale fabrication of nanoribbons remains a challenge. The graphene nanomesh, also known as a graphene antidot lattice, is a viable alternative [5, 6]. Here, the confinement potential is created by clusters of vacancies, i.e. nanoholes, arranged on a regular superlattice. They are prepared using block copolymer lithography. Two structural parameters, the cluster size and neck width of the superlattice of nanoholes, govern the electronic transport properties observed in experiments [6]. A larger energy gap is induced when both these two parameters are reduced.

Another variant of nanomesh, formed by periodic pattern of hydrogen clusters, has recently been observed for graphene grown on an iridium substrate [7]. The periodicity is due to the fact that the composite structure of graphene and iridium forms a superlattice, with the hydrogenation occurring preferentially on specific superlattice sites. The resulting structure can be regarded as a variant of the nanomesh since regions of hydrogenated graphene are highly insulating [8–10]. Nanomesh via patterned hydrogenation is a promising approach since its cluster size and neck width can be much smaller than the lithographically defined case. Indeed, the opening of a substantial energy gap has been revealed by angle-resolved photoemission spectroscopy (ARPES) [7].

Theoretical investigations of the electronic properties of nanomesh/pattern-hydrogenated graphene have been limited so far to band-structure calculations using primitive supercells [5, 7, 11, 12]. This approach can treat disorder in the cluster shape only within the same supercell [7]. Instead, in this work, we present a modeling study of pattern-hydrogenated graphene that also includes disorder across different supercells. Through calculations of the momentum-energy resolved density of states and its electrical conduction, we study the scaling of the energy gap on the parameters defining the patterned hydrogenation, i.e. cluster size, filling factor and neck width.

The model: A simple tight-binding model is employed to describe the composite structure of graphene with adsorbed hydrogen atoms [13, 14]. Within this model, the basis consists of a $2p_z$ orbital per carbon atom and a $1s$ orbital per hydrogen atom: the parameters describing the carbon-carbon hopping integral ($\gamma = 2.6$ eV), carbon-hydrogen hopping integral ($\gamma_H = 5.72$ eV), and hydrogen onsite energy ($\epsilon_H = 0$ eV) are taken from Ref. 14. Such a minimal model captures the essential physics of the hydrogenation effect pertinent for our study, that is the removal of the p_z orbital of the hydrogenated carbon atom from the π and π^* bands [34]. Since our purpose is to study the intrinsic properties of the nanomesh, we neglect the interaction with the iridium substrate in the TB model. As a consequence, particle-hole symmetry is preserved and the neutrality point remains located at the Dirac point.

Graphene on an iridium substrate forms a superlattice due to the mismatch between their respective lattice constants; 10×10 graphene unit cells are commensurate with 9×9 iridium unit cells [15]. The superlattice unit cell is represented in Fig. 1a. The supercell preserves the symmetry of the graphene unit cell [35], resulting in a honeycomb superlattice. Experimentally [7], it was shown that the hydrogen clusters tend to form in regions indicated by circles in Fig. 1a, where one graphene sublattice sits directly on top of iridium atoms. A hydrogenation model is developed to reproduce this preferential adsorption (see details in Supp. info). The model takes as input two parameters: a discrete quantity N_w which represents the cluster radius, and the filling factor n_c i.e. ratio between the number

*E-mail: rgrassi@arces.unibo.it

†Current address: IBM T.J. Watson Research Center, Yorktown Heights, New York 10598, USA

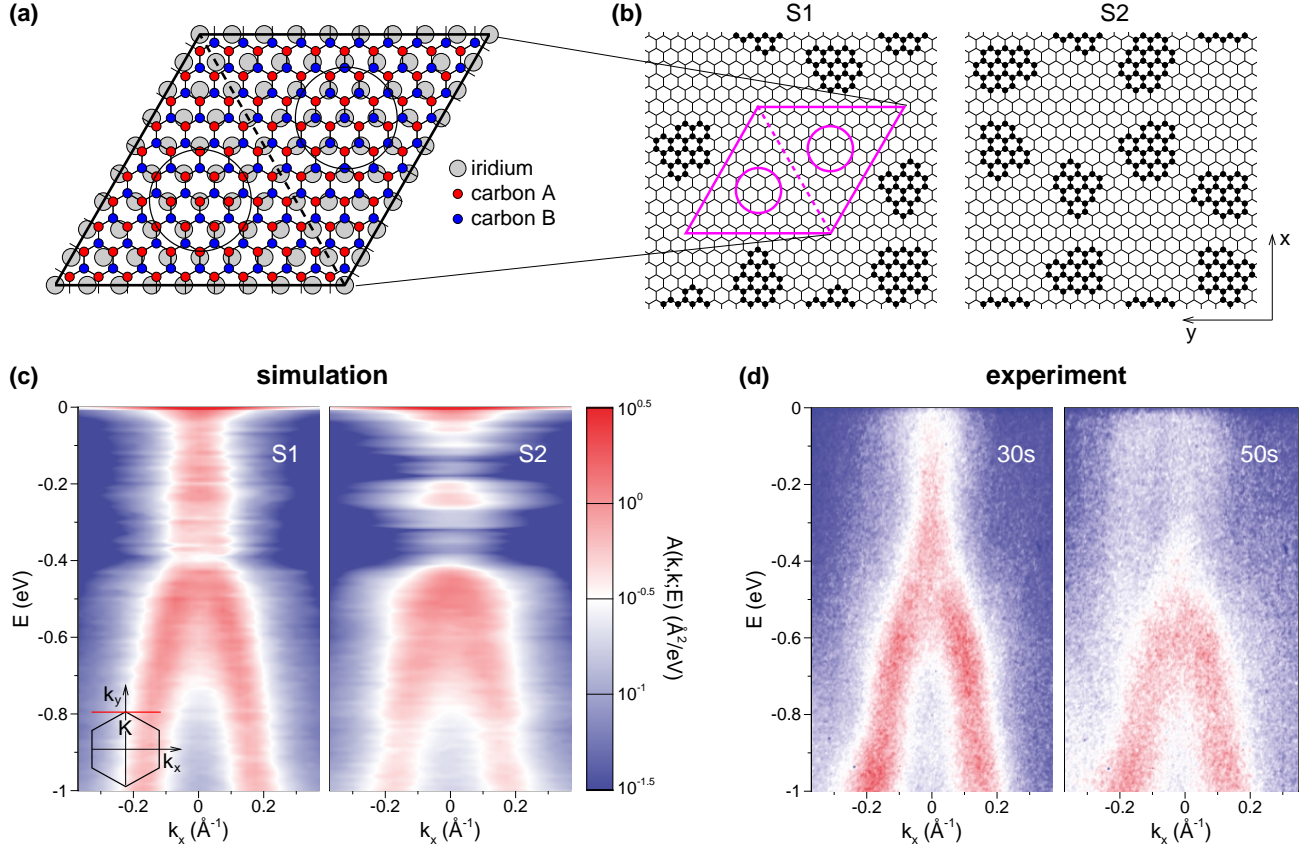


FIG. 1: Schematic representation of the atomic structure under study and comparison of the simulated k -resolved density of states in energy with the experimental ARPES. (a) Top view of the supercell of graphene on iridium substrate. The two graphene sublattices are indicated with different colors. The supercell is symmetric under reflection across the dashed line, except for the interchange of the two graphene sublattices. The two circles highlight the regions of the supercell where the clusters tend to form. (b) Top view of two hydrogenated samples with different cluster concentration. Hydrogen atoms are represented as black dots on the honeycomb graphene lattice and the iridium substrate is not shown. S1 is obtained with the model parameters $N_w = 4$, $n_c = 0.75$, while S2 with $N_w = 4$, $n_c = 1$. (c) Calculated momentum-energy resolved density of states for two sets of hydrogenated samples that correspond to the cases S1 and S2 shown in (b). 50 samples are considered for each set, the plotted quantity being the average. The inset shows the direction within the graphene Brillouin zone (red line) along which the calculation is performed. (d) Experimental ARPES intensity for different times of exposure to hydrogen (as indicated in the labels), reprinted by permission from Macmillan Publishers Ltd: Nature Materials **9**, 315, copyright 2010.

of clusters and the number of half-supercells. Two types of disorder are considered: *i*) irregular cluster edges and *ii*) a random filling of the superlattice (if $n_c < 1$). See Fig. 1b for illustrations. Each cluster is generated by adding hydrogen atoms on top of the carbon atoms belonging to a certain number of shells around the center of the half-supercell, and the edge disorder is introduced by partial hydrogenation of the outer shell. Both *A* and *B* sites are hydrogenated within each cluster, contrary to what is expected in the real structure [7]. The additional hydrogen atoms play the role of the neglected interactions with the substrate in removing the pseudo dangling bonds [7, 16].

Key features in momentum-energy resolved density of states: In order to study the electronic properties of pattern-hydrogenated graphene as seen in ARPES experiments, the calculation of the momentum-energy resolved density of states is required. This quantity is given, apart from a normalization factor, by the diagonal elements of the spectral function in momentum space, $A(\mathbf{k}, \mathbf{k}; E)$. While this quantity reduces to the usual band structure for periodic systems, it is a general concept and is valid even for disordered systems. The calculation is performed by first computing the spectral function in real space $A(\mathbf{r}, \mathbf{r}'; E)$ and then Fourier transforming to get $A(\mathbf{k}, \mathbf{k}; E)$ [36]. The calculation is done using the Green's function formalism with a recursive algorithm for periodic structures. See Supp. info for detailed numerical description and implementation.

In Fig. 1c, we plot the averaged $A(\mathbf{k}, \mathbf{k}; E)$ for two ensembles corresponding to the two realizations shown in Fig. 1b, along a path in k -space that includes the K point [37]. The convergence of the result with respect to

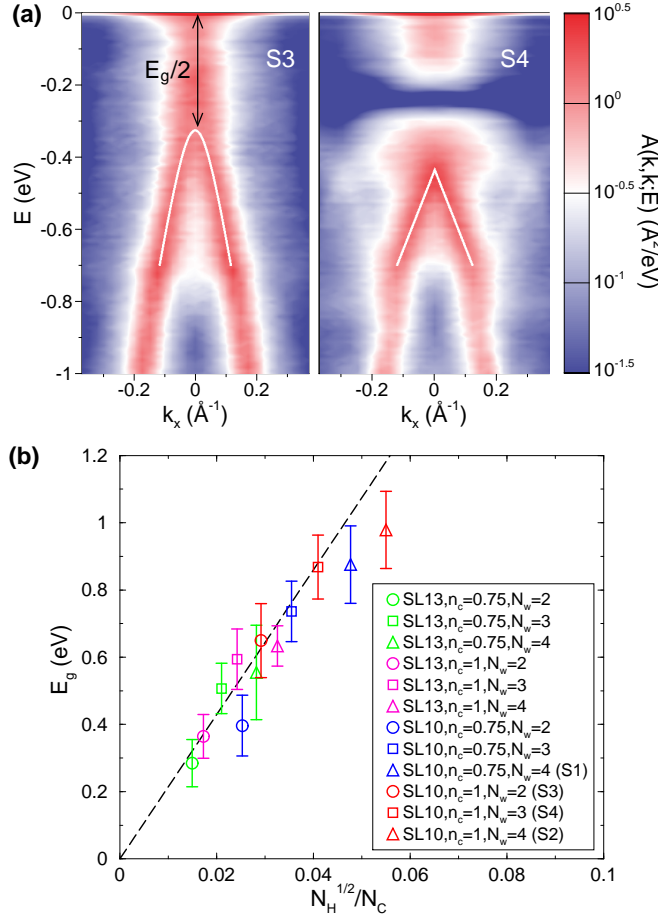


FIG. 2: Band-gap extraction. (a) Momentum-energy resolved density of states for two sets of hydrogenated samples on iridium substrate: S3 is obtained with the model parameters $N_w = 2$, $n_c = 1$, while S4 with $N_w = 3$, $n_c = 1$. Different fitting curves are used (white lines), given by Eqs. 31 and 29 of Supp. info, for S3 and S4, respectively. The band gap is extracted with respect to the fitting curve. (b) Band gap extracted for the various sets of samples and plotted as a function of $\sqrt{N_H}/N_C$, where N_H and N_C are the average number of hydrogen and carbon atoms in the half-supercell, respectively. The fitting functions used for the extraction are listed, for each set, in Table 1 of Supp. info. SL10 stands for graphene on iridium substrate (supercell made of 10×10 graphene unit cells, see Fig. 1a), while SL13 refers to graphene on a fictitious substrate (supercell made of 13×13 graphene unit cells). The error bar is a measure of the broadening of the $A(\mathbf{k}, \mathbf{k}; E)$ plot. The dashed line is the curve $E_g = 2\hbar v_F \pi / \Delta$, with $v_F = (3/2)a_{CC}|\gamma|/\hbar$ the Fermi velocity in pristine graphene and $\Delta = \sqrt{A_s/2} (N_C/\sqrt{N_H})$, where a_{CC} is the carbon-carbon distance and $A_s = a_{CC}^2 3\sqrt{3}/2$ the area of the unit cell of pristine graphene.

sample size and ensemble size has been checked, as reported in Supp. info. Only the negative energies are shown, as the conduction bands are symmetrical to the valence ones due to particle-hole symmetry. The corresponding experimental ARPES image [7] corresponding to two different hydrogen doses is shown in Fig. 1d. Several distinctive features are observed in both simulations and experiments. In both cases, the two valence branches intersect at a lower energy than the Dirac point. In addition, the signal of the states lying at the K point between $E = 0$ and the intersection energy gets suppressed with increasing hydrogen doping. These features can be interpreted as a band-gap opening. The presence of a flat band at $E = 0$ in the simulation results is a well-known effect, due to the imbalance between the two graphene sublattices [17]. The absence of these states in the experimental ARPES could be related to bond relaxation and sp^3 hybridization [18], which are neglected in the simulations.

Scaling of energy gap: The energy gap is extracted from the momentum-energy resolved density of states for different sets of samples, corresponding to different values of cluster size, filling factor, and supercell size. The supercell size is increased by considering a fictitious substrate other than iridium: for fixed cluster size, this corresponds to increasing the neck width. Fig. 2a illustrates the fitted band edges from $A(\mathbf{k}, \mathbf{k}; E)$, with details in Supp. info. An apparent universal scaling relation for the band gap is obtained when we plot the extracted

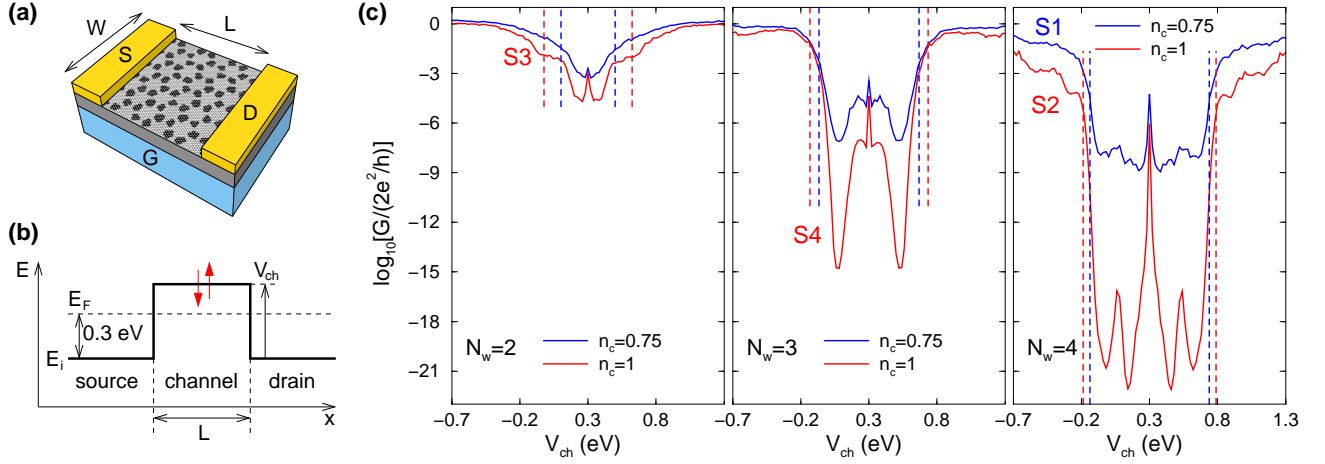


FIG. 3: Transport simulations. (a) Conceptual device under investigation: pattern-hydrogenated graphene is transferred to an insulating substrate and used as channel material of a field-effect transistor. (b) Profile of the potential energy used to simulate the structure in (a): the Fermi level in the leads E_F is kept fixed, while the barrier height V_{ch} is varied to reproduce the effect of the back gate. Pristine graphene is used for the leads. (c) Zero-temperature conductance vs. V_{ch} for various sets of samples with $W = L = 30$ nm and iridium substrate (SL10). From left to right, the cluster size, i.e. N_w , is increased; within the same plot, the cluster concentration n_c is varied. 100 samples are considered for each set and the average is done on the logarithm of the normalized conductance (a motivation for this type of averaging can be found in [22]). The vertical lines indicate the band gap from Fig. 2b.

band-gap values E_g (together with a measure of the broadening of each $A(\mathbf{k}, \mathbf{k}; E)$ plot as error bar) against the quantity $\sqrt{N_H}/N_C$, where N_H and N_C are the average number of hydrogen and carbon atoms in the half-supercell (Fig. 2b) [38]. A similar relation also applies for the case of triangular graphene nanomesh [5]. In Ref. 11, it was stated that a universal relation does not hold for honeycomb graphene nanomesh. However, Fig. 2b suggests that when disorder is included in the simulations, the scaling law $E_g = c\sqrt{N_H}/N_C$, with c a constant, can be valid at low defect coverage for honeycomb superlattices as well. This is similar to the case of graphene nanoribbons, where theoretically the band gap depends on the precise atomic configuration [19], while a general law $E_g \propto 1/W$, with W the ribbon width, is always observed in the experiments [3] and commonly attributed to disorder [20, 21]. Regarding the proportionality constant, we found that the expression $c = 2\hbar v_F \pi / \sqrt{A_s/2}$, with v_F the graphene Fermi velocity and A_s the area of the graphene unit cell, fits fairly well the numerical data (dashed line in Fig. 2b). We note that, by defining $\Delta = \sqrt{A_s/2} (N_C/\sqrt{N_H})$, the scaling relation takes the form $E_g = 2\hbar v_F \pi / \Delta$. This equation can be thought of as arising from the quantization of the graphene dispersion relation $E = \pm \hbar v_F |\mathbf{k}|$ (\mathbf{k} is here the wave vector around the K point) with $|\mathbf{k}| = \pi/\Delta$ (1D quantization in random directions): Δ can thus be interpreted as an effective confinement length.

Transport gap: Next, we examine the electronic transport properties of pattern-hydrogenated graphene. Techniques for the transfer of graphene grown on metal surfaces to an insulating substrate have recently been developed [23]. We consider a three-terminal structure as shown in Fig. 3a and aim at predicting its low-temperature, low-bias conductance. Fig. 3b illustrates the potential energy along the device. The potential energy in the source and drain leads, as a result of metal induced doping, is kept fixed with respect to the Fermi level E_F . The channel potential V_{ch} is modulated by the back gate. Graphene is aligned with its armchair direction along the longitudinal direction of the device, in order to avoid edge transport effects. Only the channel is hydrogenated while the leads are pristine graphene. The conductance is computed by using the standard Green's function technique [24] combined with a modified version of the algorithm described in Ref. 25, which is commonly used for the calculation of the lead self-energies (see Supp. info).

Fig. 3c shows the simulated, ensemble averaged zero-temperature conductance G vs. V_{ch} . The device size is kept fixed at $W = L = 30$ nm, while different sets of hydrogenated samples are considered. It can be seen that patterned hydrogenation leads to a clear transport gap, increasing with N_w and n_c . Also, the transport simulations agree well with our band-structure results: the transport gap matches the band gap from the $A(\mathbf{k}, \mathbf{k}; E)$ fitting (as indicated by vertical lines in Fig. 3c) and the peaks in the transport gap region correspond to the gap states in $A(\mathbf{k}, \mathbf{k}; E)$ [39]. The G vs. V_{ch} curve appears symmetrical, unlike the case for pristine graphene [26]. This suggests that scattering is dominated by the channel, instead of the tunneling resistance due to pn junctions.

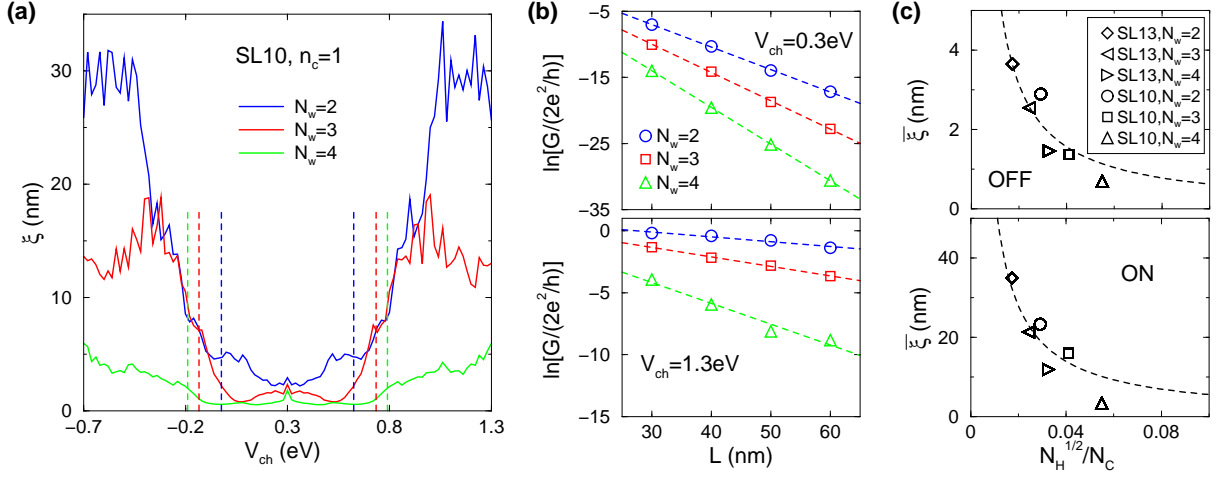


FIG. 4: Decay length extraction. (a) Decay length vs. V_{ch} for sets of samples with different cluster size and fixed supercell size (SL10) and cluster concentration ($n_c = 1$). The vertical lines indicate the band gap from Fig. 2b. (b) Example of the decay length extraction at two different V_{ch} points. The dashed lines indicate the fitting with the formula $\ln[G/(2e^2/h)] = \ln g_0 - L/\xi$. (c) Average value of the decay length in the “off” and “on” state for various sets of samples with different supercell and cluster sizes, plotted as a function of $\sqrt{N_H}/N_C$. The dashed lines indicate the fitting curve $\bar{\xi} \propto N_C/\sqrt{N_H}$. The off state is defined as the bias region $|V_{ch} - E_F| < E_G/2 - B$, where B is the half-broadening from Fig. 2b, while the on state as $0.65 \text{ eV} < |V_{ch} - E_F| < 0.75 \text{ eV}$.

Scaling of transport coefficients: Finally, we examine how G scales with L . Here, we consider devices with filling factor $n_c = 1$. The conductance is found to scale as $G \propto \exp(-L/\xi)$, where ξ is a decay length. Fig. 4a plots the extracted ξ as function of V_{ch} bias, while Fig. 4b illustrates the extraction of ξ for two particular V_{ch} values. Next, we extract the average value $\bar{\xi}$ of the decay length in the “off” and “on” state and plot it against $\sqrt{N_H}/N_C$ as we have done previously for the band gap (Fig. 4c, see caption for the definition of the “on” and “off” states). One observes that, for almost all the samples, the value of the decay length in the “off” state is about an order of magnitude smaller than the corresponding value in the “on” state. For both cases, the average decay length seems to follow the general scaling law $\bar{\xi} \propto N_C/\sqrt{N_H}$ at low to moderate hydrogenation concentrations, albeit the “off” state exhibits a smaller proportionality constant. We note that in the “off” state the exponential decrease of G with L can be explained as evanescent transport through a clean band gap [20]; the scaling law $\bar{\xi} \propto \Delta \propto 1/E_g$ is in agreement with this interpretation. In the “on” state instead, the exponential decrease of G with L is an effect of quantum localization due to disorder and ξ takes the meaning of a localization length [27]. The scaling law $\bar{\xi} \propto \Delta$ is here less clear and an exponential dependence could also be possible, as suggested by recent experiments [28]. Moreover, dephasing effects, which are ignored in our simulations, could restore a diffusive transport regime, $G \propto 1/L$, as the temperature is raised.

Conclusions: In conclusion, a simple model for pattern-hydrogenated graphene was presented. Similar features are observed in the calculated k -resolved density of states in energy and in the experimental ARPES. The scaling of the energy gap on the parameters N_C and N_H was presented, including its electronic transport properties at low temperature. Our results indicate that pattern-hydrogenated graphene is a promising approach to the engineering of graphene nanomeshes with extremely scaled cluster sizes and neck widths.

Acknowledgement: The authors gratefully acknowledge support from Network for Computational Nanotechnology for computational services.

Appendix A: Triangular versus honeycomb superlattice

While patterned hydrogenation has been experimentally demonstrated only for graphene on iridium, other substrate materials that can accommodate a monolayer graphene on their surface, e.g. rutenium [29], could in principle be used. The resulting superlattice can be of the triangular or honeycomb type. Let m and n be the size of the supercell in units of the graphene and substrate lattice constants, respectively. Here it is demonstrated that a honeycomb superlattice is obtained whenever $m/n = (3M + 1)/(3N)$ with $M, N \in \mathbb{Z}^+$.

Figs. 5a and 5b show the unit cell for the honeycomb graphene lattice and the typical triangular substrate surface layer, respectively. According to the definition of m and n given above, we assume that an $m \times m$ graphene supercell is commensurate with an $n \times n$ substrate supercell. Therefore, denoting the lattice vectors of graphene, substrate, and superlattice by \mathbf{a}_j , \mathbf{b}_j , and \mathbf{c}_j ($j = 1, 2$), respectively, we have

$$\mathbf{c}_j = m\mathbf{a}_j = n\mathbf{b}_j. \quad (\text{A1})$$

Each pair (m, n) and its multiples map to a unique composite system, e.g. $(4, 3)$ is the same as $(8, 6)$. Hence, we consider only the case where m and n are prime to each other, so that \mathbf{c}_j are the primitive lattice vectors of the superlattice. We further assume that, at some point O inside the supercell, a carbon atom sits directly on top of a substrate atom: such arrangement was found to be an energetically stable configuration [15]. We indicate the superlattice points that are equivalent to O as S_A .

As shown by Fig. 5c, in order to generate a honeycomb superlattice, there must be another point S_B inside the supercell where a carbon atom belonging to the opposite sublattice sits on top of a substrate atom. Also, for the superlattice to be regular, it can be derived that the distance between S_A and S_B must be equal to $|\mathbf{c}_1 + \mathbf{c}_2|/3$. Inspecting the graphene lattice tells us that S_A and S_B have to be separated by a distance of $(3M + 1)a_{CC}$, where $M \in \mathbb{Z}^+$ and a_{CC} is the carbon-carbon distance. Hence, we can write

$$\frac{1}{3} |\mathbf{c}_1 + \mathbf{c}_2| = \frac{m}{3} |\mathbf{a}_1 + \mathbf{a}_2| = (3M + 1)a_{CC} \quad \Rightarrow \quad m = 3M + 1, \quad M \in \mathbb{Z}^+. \quad (\text{A2})$$

In a similar fashion for the substrate, the S_A and S_B have to be separated by a distance of $N a_{SS}$, where $N \in \mathbb{Z}^+$ and a_{SS} is the interatomic distance of the substrate. We can then write

$$\frac{1}{3} |\mathbf{c}_1 + \mathbf{c}_2| = \frac{n}{3} |\mathbf{b}_1 + \mathbf{b}_2| = N a_{SS} \quad \Rightarrow \quad n = 3N, \quad N \in \mathbb{Z}^+. \quad (\text{A3})$$

In conclusion, the superlattice with the similar honeycomb structure as graphene, shown in Fig. 5c, can be generated by satisfying Eqs. A2-A3. Otherwise, the superlattice would produce a triangular structure instead, with only repeated units of S_A . Examples of honeycomb superlattices are the cases $m/n = 10/9$ (graphene on iridium, also indicated as SL10 in the manuscript) and $m/n = 13/12$ (indicated as SL13 in the manuscript).

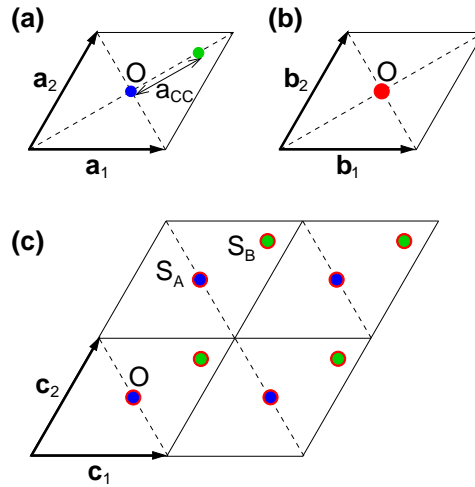


FIG. 5: (a) Graphene unit cell: the two carbon atoms are indicated with different colors. (b) Substrate unit cell. (c) Honeycomb superlattice generated by the superposition of the graphene and substrate lattices. S_A (S_B) is the point inside the supercell where a carbon atom of the A (B) graphene sublattice sits on top of a substrate atom.

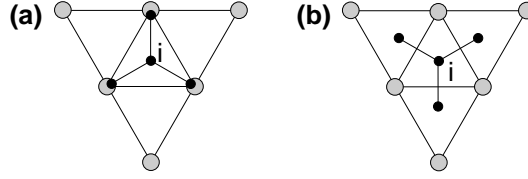


FIG. 6: Best (a) and worst (b) cases for the probability of hydrogenation of a carbon atom of index i located at a distance $d_i = a_{SS}/\sqrt{3}$ from the nearest-neighbor substrate atoms. Carbon (substrate) atoms are represented with black (gray) balls.

Appendix B: Model for patterned hydrogenation

The model for generating the hydrogen clusters is described below.

Given a sample of pristine graphene on a certain substrate, we divide the structure in supercells, the supercell subdivision being chosen so that each supercell contains one S_A and one S_B point in symmetric positions (see Fig. 1a of the main text for an illustration: the centers of the two circles correspond to an S_A or S_B point). Each carbon atom can be denoted by the pair of indexes (l, i) , where l is the index of the half-supercell to which it belongs and i is the atom index inside the whole supercell. We introduce a binary random variable $Z_{l,i} \in \{0, 1\}$ for each carbon atom: the atom is hydrogenated when $Z_{l,i} = 1$. $Z_{l,i}$ in turn is written as the product of other two binary random variables X_l and Y_i , whose probability distributions are given below.

Y_i , controls the cluster formation inside each supercell. We propose the following formula for the probability $P(Y_i = 1)$:

$$P(Y_i = 1) \equiv f(w_i), \quad f(w_i = 0) = 0, \quad \frac{df}{dw_i} \geq 0, \quad (\text{B1})$$

w_i being a quantity defined for each carbon atom as

$$w_i = \frac{d_i \left| d_i - \frac{1}{3} \sum_{\langle j \rangle} d_j - c \right|}{a_{CC}^2}, \quad (\text{B2})$$

where d_i is the xy -plane distance between the carbon atom of index i and its nearest-neighbor substrate atom (let \mathbf{r} be the position vector in the xy -plane parallel to the surface),

$$d_i = \min_k |\mathbf{r}_{C_i} - \mathbf{r}_{S_k}|, \quad (\text{B3})$$

the summation over j is restricted to the three carbon atoms that are nearest neighbor to the carbon atom of index i , and c is simply a constant,

$$c = \frac{a_{SS}}{\sqrt{3}} - \sqrt{\frac{a_{SS}^2}{3} + a_{CC}^2} - \frac{a_{SS}a_{CC}}{\sqrt{3}}. \quad (\text{B4})$$

Eqs. B1–B2 can be justified by the following considerations. Experimentally, the hydrogen clusters tend to form around the regions where one graphene sublattice is located on top of substrate atoms, while the other sublattice is far from substrate atoms and can bind to hydrogen atoms on the opposite face [7]. This translates in two conditions for the generic carbon atom to be hydrogenated. First, it should be located in between substrate lattice sites. The probability for adsorption would then increase as its distance d_i from the nearest-neighbor substrate atom increases. This effect is captured by the prefactor in (B2). However, if the considered carbon atom is located at the maximum distance from substrate atoms, equal to $a_{SS}/\sqrt{3}$, the probability for adsorption should distinguish between the case in which the three nearest-neighbor carbon atoms are located close to substrate atoms (high probability, Fig. 6a) and the case in which also the three nearest neighbors are between substrate atoms (low probability, Fig. 6b). We can note that in the first case $d_i \gg \frac{1}{3} \sum_{\langle j \rangle} d_j$, while in the second case $d_i \sim \frac{1}{3} \sum_{\langle j \rangle} d_j$. This explains the second factor in (B2), where the constant c serves only to set the probability to 0 for the worst case (Fig. 6b).

In (B1) we have omitted the actual functional dependence of $P(Y_i = 1)$ on w_i . Since this relationship depends on the physical hydrogenation process and it is unknown, we choose here a simple cut-off model. For a given superlattice unit cell, all the possible values of w_i are computed and labeled in decreasing order as $w^{(1)}, w^{(2)}, \dots$ (the location of

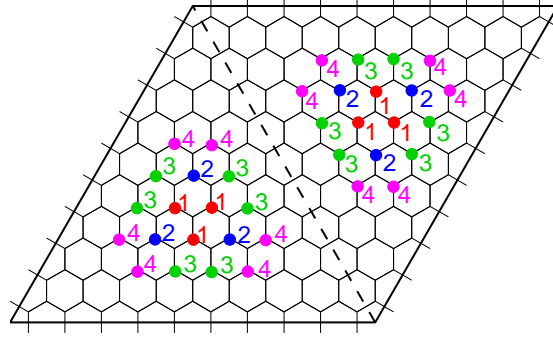


FIG. 7: Supercell of graphene on iridium: location of the carbon atoms with the four largest values of w_i , i.e. $w_i = w^{(1)}$, $w_i = w^{(2)}$, etc.

the corresponding carbon atoms is shown in Fig. 7 for the case of iridium substrate). Then, the probability $P(Y_i = 1)$ is assigned as

$$P(Y_i = 1) \equiv f(w_i) = \begin{cases} 1 & \text{if } w_i = w^{(j)} \text{ with } j < N_w, \\ 0.5 & \text{if } w_i = w^{(j)} \text{ with } j = N_w, \\ 0 & \text{if } w_i = w^{(j)} \text{ with } j > N_w. \end{cases} \quad (\text{B5})$$

With this method, a cluster of hydrogen atoms is formed around the sites where the quantity w_i tends to grow (i.e. around S_A and S_B). The input parameter N_w controls the size of this cluster. The disorder is only located at the cluster edges.

X_l , instead, is used to generate the superlattice disorder, which consists in some hydrogen clusters being randomly missing from the superlattice. The probability $P(X_l = 1)$ is set equal to the input parameter n_c , with $0 \leq n_c \leq 1$, which therefore assumes the meaning of the ratio between the average number of hydrogenated half-supercells (or equivalently average number of clusters) and the total number of half-supercells.

The hydrogenation model described above produces clusters inside which only one graphene sublattice is hydrogenated. This leads to the formation of midgap states in the electronic structure, associated with dangling bonds. However, these states are believed to be an artefact of the TB model, due to the fact that bond relaxation is neglected. To avoid this, after hydrogen atoms are generated according to the method described above, a final step is introduced: additional hydrogen atoms are placed on top of the carbon atoms that have at least two nearest neighbors being hydrogenated.

Appendix C: Calculation of momentum-energy resolved density of states

We consider a sample composed of N_1 and N_2 graphene unit cells along the directions of \mathbf{a}_1 and \mathbf{a}_2 , respectively, with periodic boundary conditions on both directions. N_1 and N_2 are chosen to be multiples of m , the size of the supercell in units of the graphene lattice constant, so that the sample contains exactly $\frac{N_1}{m} \times \frac{N_2}{m}$ supercells. The sample is then hydrogenated as described in the previous section. An example, obtained with $N_1 = N_2 = 20$, $m/n = 10/9$ (graphene on iridium), $n_c = 0.75$, and $N_w = 4$, is shown in Fig. 8a. The actual samples that are simulated are much larger: $N_1 = N_2 = 120$ is used for the SL10 case, while $N_1 = N_2 = 117$ for the SL13 case. Such values have been checked to be large enough to ensure the convergence of $A(\mathbf{k}, \mathbf{k}; E)$ (see next section).

The generic orbital of the TB representation can be indicated as $|\mathbf{l}, q\rangle$, where \mathbf{l} is the graphene lattice vector, i.e. $\mathbf{l} = \sum_i n_i \mathbf{a}_i$ with $n_i = 1, \dots, N_i$, and q is the orbital index within the cell ($q = 1, 2$ for the A and B carbon orbitals, respectively). This is the real space representation. However, one could also work in the k -space representation, which is obtained by restricting $q = 1, 2$ (projected space for carbon atoms only) and by using as basis the set $\{|\mathbf{k}, q\rangle\}$ defined by $\langle \mathbf{l}, q_1 | \mathbf{k}, q_2 \rangle = \delta_{q_1, q_2} e^{i\mathbf{k} \cdot \mathbf{l}} / \sqrt{N_1 N_2}$. The \mathbf{k} vectors are discrete because a finite volume is considered: if \mathbf{d}_i are the primitive vectors of the reciprocal lattice, i.e. $\mathbf{a}_i \cdot \mathbf{d}_j = 2\pi \delta_{i,j}$, we have $\mathbf{k} = (m_1/N_1)\mathbf{d}_1 + (m_2/N_2)\mathbf{d}_2$, where $m_1, m_2 \in \mathbb{Z}$ (it can be proved that the K point is included in the grid if both N_1 and N_2 are multiple of 3). Also, since the set of \mathbf{l} vectors is discrete, it follows that only a number $N_1 N_2$ of \mathbf{k} vectors, spanning a Brillouin zone, give rise to independent basis vectors.

Let H be the electron Hamiltonian. We recall that the spectral function at the energy E is defined as $A = i(G^r - G^a)$, where $G^r = [(E + i\eta)I - H]^{-1}$ is the retarded Green's function and $G^a = G^{r\dagger}$ is the advanced one (η is an infinitesimal

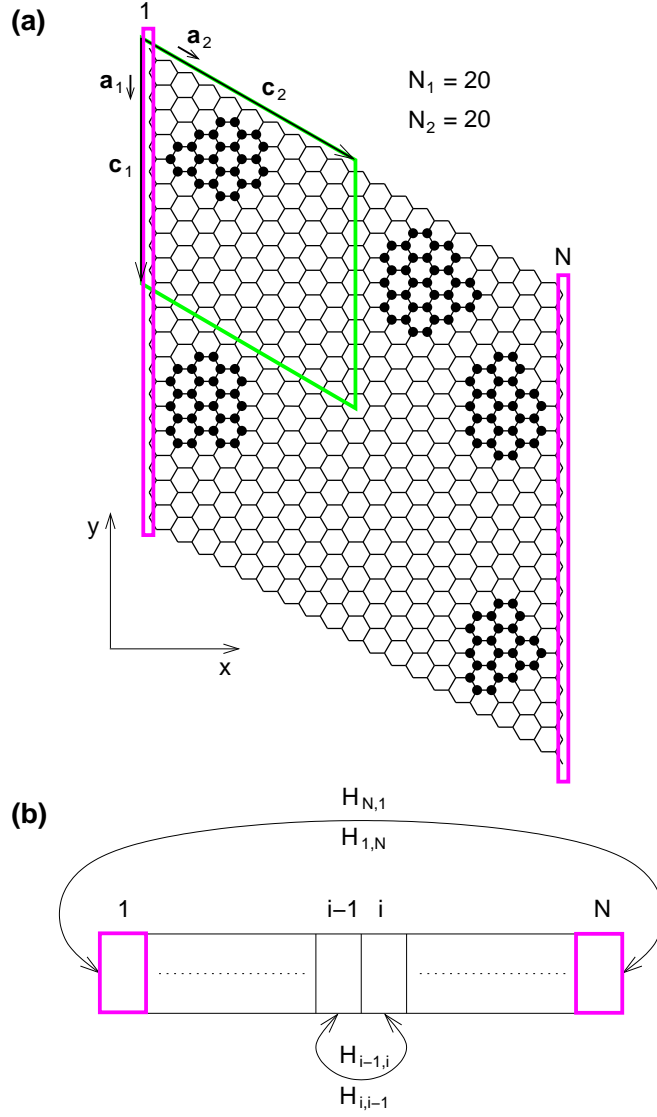


FIG. 8: (a) Example of sample used for calculating the momentum-energy resolved density of states. The structure is made of $N_1 \times N_2$ graphene unit cells, with N_1 and N_2 being chosen so that the sample is a multiple of the superlattice unit cell (region enclosed by the green line). Periodic boundary conditions are applied on both the \mathbf{a}_1 and \mathbf{a}_2 directions. (b) The same structure can be viewed as a linear chain of $N = 2N_2$ slabs (each slab corresponding to a row of atoms) with a periodic closure at the two ends.

positive quantity). While the diagonal elements of A in real space have the meaning of a density of states in energy and physical space, its diagonal elements in k -space give the density of states in energy and momentum, which in turn corresponds to the physical quantity measured by ARPES. The diagonal elements of A in k -space can be expanded as

$$\begin{aligned}
 A(\mathbf{k}, q; \mathbf{k}, q; E) &= \sum_{\mathbf{l}_1, q_1} \sum_{\mathbf{l}_2, q_2} \langle \mathbf{k}, q | \mathbf{l}_1, q_1 \rangle A(\mathbf{l}_1, q_1; \mathbf{l}_2, q_2; E) \langle \mathbf{l}_2, q_2 | \mathbf{k}, q \rangle \\
 &= \frac{1}{N_1 N_2} \sum_{\mathbf{l}_1} \sum_{\mathbf{l}_2} e^{-i\mathbf{k} \cdot \mathbf{l}} A(\mathbf{l}_1, q; \mathbf{l}_1 - \mathbf{l}, q; E),
 \end{aligned} \tag{C1}$$

where we recognize a discrete Fourier transform with respect to the relative variable $\mathbf{l} = \mathbf{l}_1 - \mathbf{l}_2$. We define

$$A(\mathbf{k}, \mathbf{k}; E) = \frac{A_s}{(2\pi)^2} \sum_q A(\mathbf{k}, q; \mathbf{k}, q; E), \tag{C2}$$

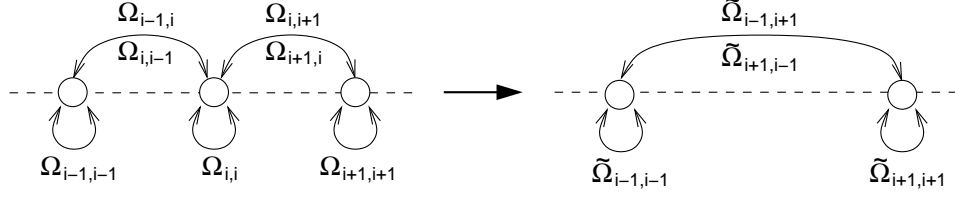


FIG. 9: Pictorial representation of the renormalization method in [30]: the circles represent nodes (or slabs) and the arrows the different types of coupling between them. The method consists in eliminating node i , while leaving unchanged the solution for all the remaining nodes by a proper renormalization of the matrix blocks of Ω .

where $A_s = a_{\text{CC}}^2 3\sqrt{3}/2$ is the area of the graphene unit cell, so that $A(\mathbf{k}, \mathbf{k}; E)/(2\pi)$ gives the number of states per unit energy, per unit \mathbf{k} , and per graphene unit cell (apart from spin degeneracy). The calculation of $A(\mathbf{k}, \mathbf{k}; E)$ is thus performed by first computing the spectral function in real space and then Fourier transforming according to (C1-C2).

A method to compute A in real space is by diagonalization of H . Indeed, if $\{|\psi_\alpha\rangle\}$ are the orthonormal eigenstates of H with corresponding eigenvalues $\{\epsilon_\alpha\}$, we have

$$A(\mathbf{l}_1, q_1; \mathbf{l}_2, q_2; E) = \sum_{\alpha} \frac{2\eta}{(E - \epsilon_\alpha)^2 + \eta^2} \psi_\alpha(\mathbf{l}_1, q_1) \psi_\alpha^*(\mathbf{l}_2, q_2) \xrightarrow{\eta \rightarrow 0} 2\pi \sum_{\alpha} \delta(E - \epsilon_\alpha) \psi_\alpha(\mathbf{l}_1, q_1) \psi_\alpha^*(\mathbf{l}_2, q_2), \quad (\text{C3})$$

with $\psi_\alpha(\mathbf{l}, q) = \langle \mathbf{l}, q | \psi_\alpha \rangle$ the generic eigenfunction in real space. The numerical computation of (C3) can be efficiently done by setting a finite value of η and by finding for each energy E the eigenvalues (and corresponding eigenvectors) that are closest to it, using well known methods for large and sparse eigenvalue problems [40]. Nevertheless, we propose here an alternative method based on Green's functions. Although in the case considered here our method does not improve the computational time with respect to the diagonalization technique, because almost all the off-diagonal of the spectral function in real space need to be calculated, the method could be interesting in other situations, such as for the calculation of the local density of states in real space, where only few elements of the spectral function are needed. The method could also be useful when the storage of the eigenvectors is a major problem.

As illustrated in Fig. 8a, we divide the sample in $N = 2N_2$ slabs along the \mathbf{a}_2 direction so that each slab corresponds to a row of atoms (the choice between \mathbf{a}_1 and \mathbf{a}_2 is arbitrary). The structure is therefore of the type in Fig. 8b: a linear chain of slabs with a periodic boundary conditions at the two ends. Using block matrix notation in real space, $\Omega = (E + i\eta)I - H$ has the form

$$\Omega = \begin{pmatrix} \Omega_{1,1} & \Omega_{1,2} & & & \Omega_{1,N} \\ \Omega_{2,1} & \Omega_{2,2} & \Omega_{2,3} & & \\ & \Omega_{3,2} & \ddots & \ddots & \\ & & \ddots & \ddots & \Omega_{N-1,N} \\ \Omega_{N,1} & & & \Omega_{N,N-1} & \Omega_{N,N} \end{pmatrix}, \quad (\text{C4})$$

where each block represents the coupling between a pair of slabs. We notice that only the elements of the spectral function that connect orbitals belonging to the same graphene sublattice (i.e. same q) are needed in (C1-C2). From Fig. 8a, it can be seen that there is a correspondence between $q = 1, 2$ and the odd and even slabs, respectively. Therefore, only the matrix block of A (and thus of G^r) connecting slabs with the same parity need to be calculated. In order to avoid the calculation of the unnecessary matrix blocks, the renormalization method [30] is employed: an equivalent Ω^{odd} (Ω^{even}) matrix for the odd (even) slabs alone is computed by decimating the even (odd) ones. We recall that the decimation of a single node i from a linear chain, as depicted in Fig. 9, is achieved by renormalizing the matrix blocks of Ω for the adjacent nodes according to the formulas [30]:

$$\begin{aligned} \tilde{\Omega}_{i-1,i-1} &= \Omega_{i-1,i-1} - \Omega_{i-1,i} (\Omega_{i,i})^{-1} \Omega_{i,i-1}, \\ \tilde{\Omega}_{i-1,i+1} &= -\Omega_{i-1,i} (\Omega_{i,i})^{-1} \Omega_{i,i+1}, \\ \tilde{\Omega}_{i+1,i-1} &= -\Omega_{i+1,i} (\Omega_{i,i})^{-1} \Omega_{i,i-1}, \\ \tilde{\Omega}_{i+1,i+1} &= \Omega_{i+1,i+1} - \Omega_{i+1,i} (\Omega_{i,i})^{-1} \Omega_{i,i+1}. \end{aligned} \quad (\text{C5})$$

By repeated use of (C5), the following algorithm can be derived to compute the renormalized matrix Ω^{odd} for the odd slabs alone (Fig. 10):

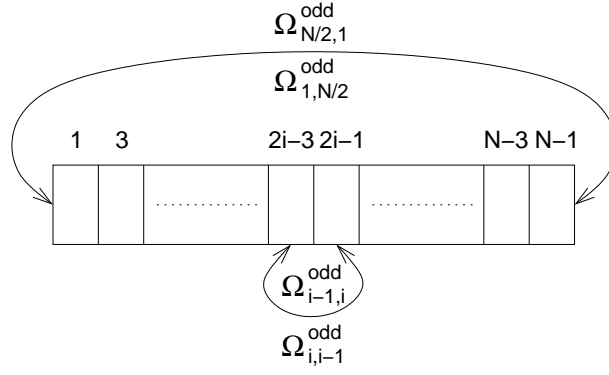
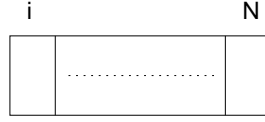


FIG. 10: Same structure as in Fig. 8b, where the even slabs have been decimated.

FIG. 11: Structure corresponding to $g^{r\mathcal{R},(i)}$.

1. for $i = 1, \dots, N/2$ initialize

$$\Omega_{i,i}^{\text{odd}} = \Omega_{2i-1,2i-1}; \quad (\text{C6})$$

2. for $i = 1, \dots, N/2$

- set $j = 2i$, $k = \text{mod}(i, \frac{N}{2}) + 1$, $l = \text{mod}(j, N) + 1$
- compute

$$\begin{aligned} \Omega_{i,i}^{\text{odd}} &= \Omega_{i,i}^{\text{odd}} - \Omega_{j-1,j}(\Omega_{j,j})^{-1}\Omega_{j,j-1}, \\ \Omega_{k,k}^{\text{odd}} &= \Omega_{k,k}^{\text{odd}} - \Omega_{l,j}(\Omega_{j,j})^{-1}\Omega_{j,l}, \\ \Omega_{i,k}^{\text{odd}} &= -\Omega_{j-1,j}(\Omega_{j,j})^{-1}\Omega_{j,l}, \\ \Omega_{k,i}^{\text{odd}} &= -\Omega_{l,j}(\Omega_{j,j})^{-1}\Omega_{j,j-1}. \end{aligned} \quad (\text{C7})$$

A similar algorithm can be derived for Ω^{even} . Both Ω^{odd} and Ω^{even} have the same shape as Ω (Eq. C4), but with $N \rightarrow N/2$. In the following, we will therefore refer to Ω for brevity, implicitly assuming that what we say must be applied separately to Ω^{odd} and Ω^{even} .

The retarded Green's function at the energy E is obtained by inverting Ω . Here we present an algorithm for recursively calculating the blocks of G^r , extending the one in [31] for the case of $\Omega_{1,N}, \Omega_{N,1} \neq 0$. The algorithm is based on Dyson's equations. We recall that if the Hamiltonian is expressed as $H = H_0 + H_1$ (H_0 is called the unperturbed Hamiltonian and H_1 the perturbation one), the Dyson equations are given by $G^r = G_0^r + G_0^r H_1 G^r = G_0^r + G^r H_1 G_0^r$, where G_0^r is the retarded Green's function corresponding to H_0 . While the formulas that are presented here directly exploit time reversal symmetry, i.e. the fact that $G^r = (G^r)^T$, their extension to the general case is straightforward.

The first part of the algorithm consists in the calculation of certain blocks of $g^{r\mathcal{R},(i)}$ for $i = 1, \dots, N$, where $g^{r\mathcal{R},(i)}$ is the retarded Green's function corresponding to the structure composed of only the nodes from i to N without the periodic closure (Fig. 11). Indeed, by applying Dyson's equations to the structure in Fig. 11 with the coupling blocks $-\Omega_{i,i+1}$ and $-\Omega_{i+1,i}$ treated as the perturbation Hamiltonian, it is possible to relate $g^{r\mathcal{R},(i)}$ to $g^{r\mathcal{R},(i+1)}$ and derive the following equations:

1. initialize

$$g_{N,N}^{r\mathcal{R},(N)} = (\Omega_{N,N})^{-1}; \quad (\text{C8})$$

2. for $i = N - 1, \dots, 1$ compute

$$g_{i,i}^{r\mathcal{R},(i)} = (\Omega_{i,i} - \Omega_{i,i+1} g_{i+1,i+1}^{r\mathcal{R},(i+1)} \Omega_{i+1,i})^{-1}, \quad (\text{C9})$$

$$g_{i,N}^{r\mathcal{R},(i)} = -g_{i,i}^{r\mathcal{R},(i)} \Omega_{i,i+1} g_{i+1,N}^{r\mathcal{R},(i+1)}, \quad (\text{C10})$$

$$g_{N,i}^{r\mathcal{R},(i)} = (g_{i,N}^{r\mathcal{R},(i)})^T, \quad (\text{C11})$$

$$g_{N,N}^{r\mathcal{R},(i)} = g_{N,N}^{r\mathcal{R},(i+1)} - g_{N,i+1}^{r\mathcal{R},(i+1)} \Omega_{i+1,i} g_{i,N}^{r\mathcal{R},(i)}. \quad (\text{C12})$$

The blocks $g_{N,N}^{r\mathcal{R},(i)}$ for $i > 1$ can be discarded. The second part of the algorithm is obtained by applying again Dyson's equations, but to the original structure in Fig. 8b, with the perturbation Hamiltonian given by the coupling blocks $-\Omega_{i-1,i}$, $-\Omega_{i,i-1}$, $-\Omega_{N,1}$, and $-\Omega_{1,N}$. The formulas are as follows

3. initialize

$$G_{1,1}^r = \left[I + g_{1,N}^{r\mathcal{R},(1)} \Omega_{N,1} - g_{1,1}^{r\mathcal{R},(1)} \left(I + \Omega_{1,N} g_{N,1}^{r\mathcal{R},(1)} \right)^{-1} \Omega_{1,N} g_{N,N}^{r\mathcal{R},(1)} \Omega_{N,1} \right]^{-1} \times \\ \times \left[g_{1,1}^{r\mathcal{R},(1)} - g_{1,1}^{r\mathcal{R},(1)} \left(I + \Omega_{1,N} g_{N,1}^{r\mathcal{R},(1)} \right)^{-1} \Omega_{1,N} g_{N,1}^{r\mathcal{R},(1)} \right]; \quad (\text{C13})$$

4. for $i = 1, \dots, N$

– if $i > 1$ compute

$$G_{i,i}^r = g_{i,i}^{r\mathcal{R},(i)} - g_{i,i}^{r\mathcal{R},(i)} \Omega_{i,i-1} G_{i-1,i}^r - g_{i,N}^{r\mathcal{R},(i)} \Omega_{N,1} G_{1,i}^r, \quad (\text{C14})$$

– for $j = i + 1, \dots, N$ compute

$$G_{i,j}^r = -G_{i,j-1}^r \Omega_{j-1,j} g_{j,j}^{r\mathcal{R},(j)} - G_{i,1}^r \Omega_{1,N} g_{N,j}^{r\mathcal{R},(j)}, \quad (\text{C15})$$

$$G_{j,i}^r = (G_{i,j}^r)^T. \quad (\text{C16})$$

As soon as each block of G^r is computed, its contribution to (C1) can be evaluated and the block is ready for being discarded unless is used later by the algorithm. It can be checked that the first row of G^r has to be fully saved, while $G_{i,j}^r$ is needed to calculate $G_{i,j+1}^r$ and $G_{i,i+1}^r$ to calculate $G_{i+1,i+1}^r$. In conclusion, compared to the direct inversion of the matrix in (C4), the proposed algorithm allows to reduce of about a factor of 4 the number of blocks that are calculated, due to both the decimation of nodes with different parity and the exploitation of time reversal symmetry; moreover, the blocks are recursively calculated one after the other, thus saving memory. We note that our algorithm could also be used for partial inversion of (C4), as in the case of the calculation of the density of states in energy, where only the diagonal elements of the spectral function in real space are required: in this case, for $i > 1$, (C15) can be limited to $j = i + 1$ and (C16) is not required. Regarding the stability of the overall algorithm, it should be pointed out that a value of $\eta = 10^{-3}$ eV was necessary in the simulations to avoid numerical artefacts; however, the corresponding broadening introduced in the $A(\mathbf{k}, \mathbf{k}; E)$ plot is way much smaller than the one due to disorder.

Appendix D: Convergence study w.r.t. sample size (and ensemble size)

In order to account for disorder across different cells of the superlattice, each sample has to be large enough so that it contains a sufficient number of clusters and the effect of the periodic boundary condition is washed out. In addition, the size of the sample determines the discretization step in k -space (the total number of \mathbf{k} points along the red line in the inset of Fig. 1c of the main text is equal to N_2): for a high-quality plot of $A(\mathbf{k}, \mathbf{k}; E)$, each sample has to be large enough so that the grid in k -space is sufficiently fine.

In Figs. 12, we report a comparison of $A(\mathbf{k}, \mathbf{k}; E)$ plots obtained by varying the sample size or the ensemble size. For the case with $N_1 \times N_2 = 30 \times 30$ and 20 samples (Fig. 12a), the number of \mathbf{k} points is small and the dispersion looks quite vague (also note that this is a zoomed view around the K point, so that only about 1/4 out of the N_2 points along the path are shown). Interestingly, the number of clusters seems to be already large enough to destroy the superlattice band structure: no repeated bands are visible, while the states are clearly arising from the graphene Dirac cone. The plot greatly improves when the size of the samples is increased to 90×90 (Fig. 12b). However, further increasing the sample size (Fig. 12c) or the number of samples (Fig. 12d), gives only a slight improvement, which means that the result has already well converged.

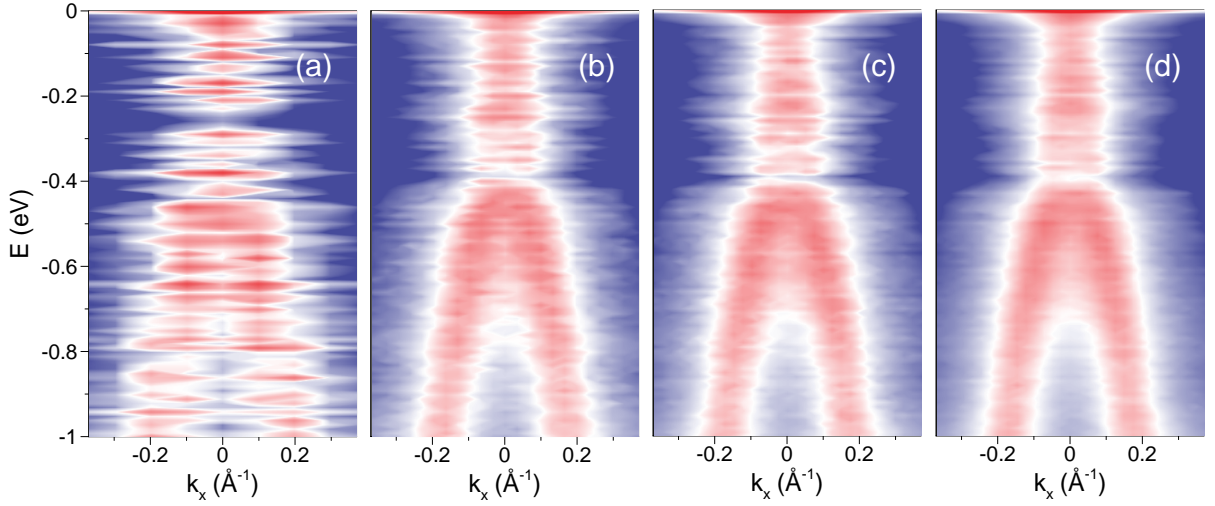


FIG. 12: Averaged $A(\mathbf{k}, \mathbf{k}; E)$ for ensembles with different sample size or ensemble size, all generated with the same supercell size (SL10), filling factor ($n_c = 0.75$), and cluster size ($N_w = 4$): (a) $N_1 \times N_2 = 30 \times 30$, 20 samples; (b) 90×90 , 20 samples; (c) 120×120 , 20 samples; (d) 120×120 , 50 samples. The color scale and path in k -space are the same as in the main text.

Appendix E: Procedure for band-gap extraction

The band gap is extracted from each (ensemble-averaged) $A(\mathbf{k}, \mathbf{k}; E)$ plot using a fitting technique. We recall that the path in k -space is the one shown in the inset of Fig. 1c in the main text, so that $\mathbf{k} = (k_x, K_y)$, where K_y is the k_y -coordinate of the K point. Since k_y is fixed, we use the simplified notation $A(k_x, k_x; E)$. The fitting procedure is composed of the following steps.

1. Manually choose a range of energies $[E_1, E_2]$ where to apply the fitting.
2. Find for each energy $E \in [E_1, E_2]$ the k_x coordinate where the intensity is maximum, separately for positive and negative k_x :

$$k_x^+(E) \quad \text{such that} \quad A(k_x^+, k_x^+; E) = \max_{k_x \geq 0} A(k_x, k_x; E), \quad (\text{E1})$$

$$k_x^-(E) \quad \text{such that} \quad A(k_x^-, k_x^-; E) = \max_{k_x \leq 0} A(k_x, k_x; E). \quad (\text{E2})$$

3. Compute for each $E \in [E_1, E_2]$ the values $w^+(E)$ and $w^-(E)$ as follows

$$w^+(E) = \frac{A(k_x^+(E), k_x^+(E); E)}{\max_{E' \in [E_1, E_2]} A(k_x^+(E), k_x^+(E); E')}, \quad (\text{E3})$$

$$w^-(E) = \frac{A(k_x^-(E), k_x^-(E); E)}{\max_{E' \in [E_1, E_2]} A(k_x^-(E), k_x^-(E); E')}. \quad (\text{E4})$$

4. Apply a least-square fitting to the set of points $\{(E, k_x^+(E))\} \cup \{(E, k_x^-(E))\}$ with $E \in [E_1, E_2]$, by using $w^+(E)$ and $w^-(E)$ as weights and one of the following dispersion relations as fitting curve:

$$E = \pm \left(\hbar v |k_x| + \frac{E_g}{2} \right), \quad (\text{E5})$$

$$E = \pm \left(\frac{\hbar^2 k_x^2}{2m} + \frac{E_g}{2} \right), \quad (\text{E6})$$

$$E = \pm \sqrt{\frac{\hbar^2 E_g k_x^2}{2m} + \left(\frac{E_g}{2} \right)^2}. \quad (\text{E7})$$

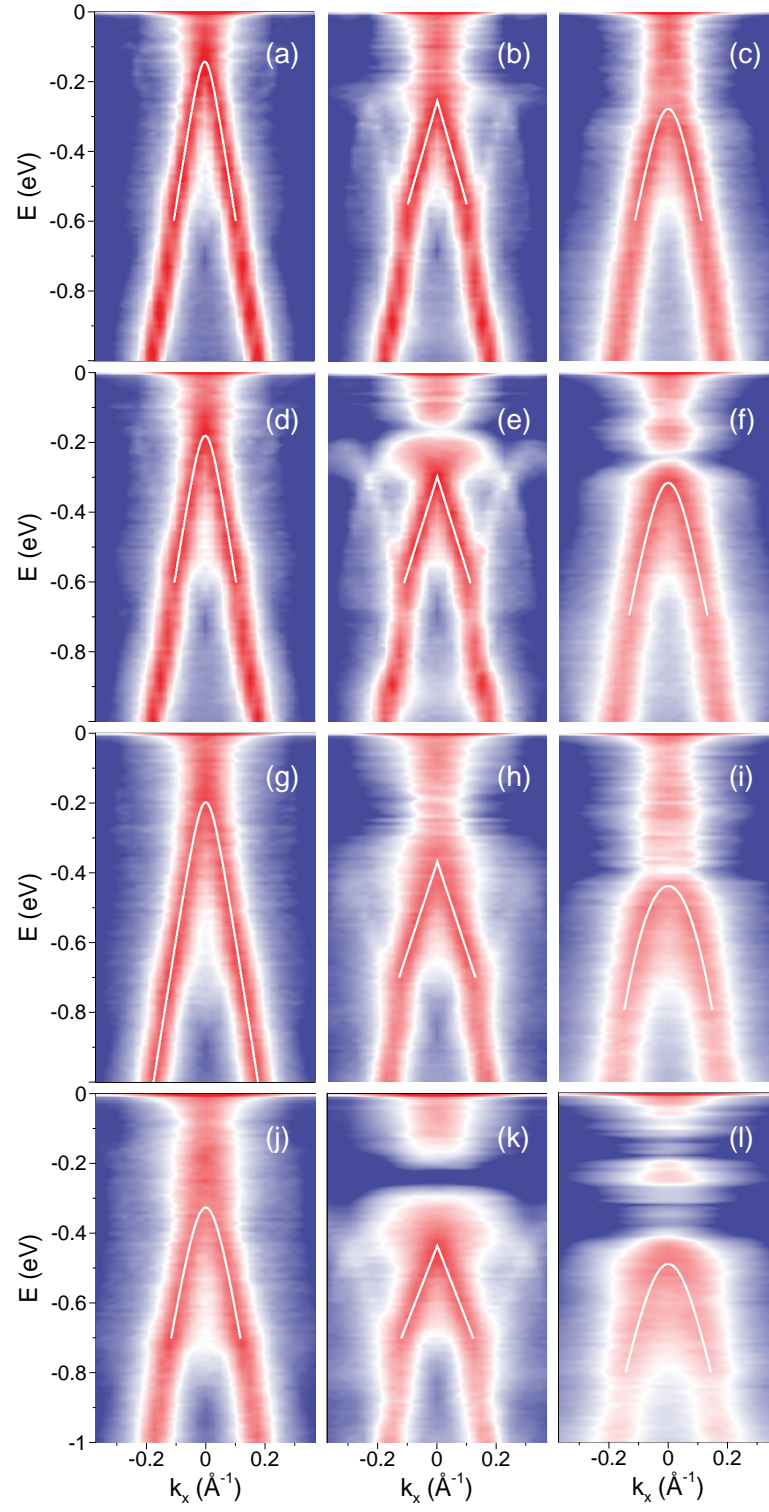


FIG. 13: Plot of averaged $A(\mathbf{k}, \mathbf{k}; E)$ and fitting curve, for all the set of samples studied in this work. (i) and (l) are the same plots as in Fig. 1c of the manuscript, (j) and (k) the same as in Fig. 2a of the manuscript.

set	SL	n_c	N_w	E_1 (eV)	E_2 (eV)	fit.	$E_g/2$ (eV)	v/v_F	m/m_0	k_x^B (\AA^{-1})	$2B$ (eV)
(a)	13	0.75	2	0.05	0.6	LP	0.285		0.035	0.101	0.14
(b)	13	0.75	3	0.2	0.55	L	0.507	0.537		0.101	0.15
(c)	13	0.75	4	0.25	0.6	LP	0.555		0.094	0.101	0.28
(d)	13	1	2	0.1	0.6	LP	0.365		0.046	0.101	0.13
(e)	13	1	3	0.3	0.6	L	0.594	0.491		0.101	0.18
(f)	13	1	4	0.25	0.7	LP	0.633		0.110	0	0.12
(g)	10	0.75	2	0.1	1	LP	0.396		0.048	0.098	0.18
(h)	10	0.75	3	0.4	0.7	L	0.736	0.465		0.098	0.18
(i)	10	0.75	4	0.4	0.8	P	0.876		0.236	0	0.23
(j)	10	1	2	0.3	0.7	LP	0.650		0.087	0.098	0.22
(k)	10	1	3	0.35	0.7	L	0.868	0.393		0.098	0.19
(l)	10	1	4	0.35	0.8	LP	0.979		0.192	0	0.23

TABLE I:
Parameters of the $A(\mathbf{k}, \mathbf{k}; E)$ fitting and broadening extraction.

The result of the fitting for negative E is shown in Fig. 13, superimposed to the original $A(\mathbf{k}, \mathbf{k}; E)$ plot.

For each $A(\mathbf{k}, \mathbf{k}; E)$ plot, a measure of the broadening is also extracted. We consider a specific k_x value, k_x^B , and compute the quantity $2B$ as the difference between the two energies at which the function $A(k_x^B, k_x^B; E)$ decreases to half of its maximum value.

The input and output parameters of the band-gap and broadening extraction are collected in Table I for each set of sample: L, P, and LP refer to the fitting curves (E5), (E6) and (E7), respectively, $v_F = (3/2)a_{CC}|\gamma|/\hbar$ is the graphene Fermi velocity, and m_0 is the electron rest mass.

The use of different fitting curves deserves an explanation. We notice that relation (E7) is the most physical one since it describes the 1D quantization of the graphene dispersion relation. However, for $N_w = 3$, i.e. cases (b), (e), (h), and (k) of Fig. 13 and Table I, the averaging effect of disorder at the cluster edges seems to be not strong enough to reach the typical dispersion relation (in fact, the repeated bands of the superlattice are still slightly visible in the $A(\mathbf{k}, \mathbf{k}; E)$ plot), and the functional dependence in (E5) gives a better fitting. Also, when the parabolicity is large in the energy range of interest, such as in cases (i) and (l) of Fig. 13 and Table I, the use of (E6) instead of (E7) does not make a significant difference in the gap value.

Appendix F: Transport calculation

We consider a structure such as the one represented in Fig. 14a. The structure is divided in slabs along the longitudinal direction so that each slab corresponds to an atomic row. The slabs inside the device region are numbered from 1 to N .

The zero temperature conductance G is given by the transmission function $T(E)$ at the Fermi energy E_F ,

$$G = \frac{2e^2}{h} T(E = E_F), \quad (\text{F1})$$

where in turn the transmission function is computed through Green's functions as [24]

$$T(E) = \text{Tr} [\Gamma^L G^r \Gamma^R G^a]. \quad (\text{F2})$$

In this equation, $\Gamma^{L/R} = i(\Sigma^{r,L/R} - \Sigma^{a,L/R})$ is the broadening function due to the left/right lead, where $\Sigma^{r,L/R}$ is the self-energy representing the renormalization of the Hamiltonian of the device region alone due to the presence of the semi-infinite left/right lead, and $\Sigma^{a,L/R} = (\Sigma^{r,L/R})^\dagger$. Since the only non-null block of $\Sigma^{r,L}$ is $\Sigma_{1,1}^{r,L}$ and the only non-null one of $\Sigma^{r,R}$ is $\Sigma_{N,N}^{r,R}$, (F2) can be rewritten as

$$T(E) = \text{Tr} [\Gamma_{1,1}^L G_{1,N}^r \Gamma_{N,N}^R G_{N,1}^a]. \quad (\text{F3})$$

The calculation of the G^r blocks can be efficiently performed using well known methods, such as the already mentioned recursive algorithm [31], or a combination of the recursive and the renormalization algorithms [32], and therefore it is not treated here. Instead, we focus on the calculation of the lead self-energies.

We notice that in our case the unit cell of each lead is made of $M = 4$ slabs (Fig. 14b). Here, we propose an algorithm, based on the renormalization method [30], to reduce the size of the unit cell to only one slab, such that the usual Sancho-Rubio algorithm [25] can then be applied on matrices having a reduced size, thus saving computational time. We notice that, in the specific case considered in this work, analytical expressions for the self-energies could have been used [33]. However, the numerical technique presented here is more general: for example, it can also be applied in the presence of a magnetic field.

We consider only the case of a left lead, the generalization to the right case being straightforward. The self-energy due to the left lead is defined as

$$\Sigma_{1,1}^{r,L} = \Omega_{1,0} g_{0,0}^r \Omega_{0,1}, \quad (\text{F4})$$

where $\Omega = (E + i\eta)I - H$ and g^r is the retarded Green's function for the case in which the coupling between the device and the leads is set to zero [24]. Suppose that the unit cell of the lead contains M slabs. The matrix Ω^L of the isolated left lead has thus the structure

$$\Omega^L = \begin{pmatrix} \ddots & & \ddots & & & \\ \ddots & & \ddots & & & \\ & \ddots & \ddots & \Omega_{-M,-M+1} & \Omega_{-M,-M+1}^C & \\ & & \Omega_{-M+1,-M} & \Omega_{-M+1,-M}^C & \Omega_{-M,-M+1} & \\ & & & \Omega_{-M+1,-M} & \Omega_{-M,-M+1}^C & \end{pmatrix}, \quad (\text{F5})$$

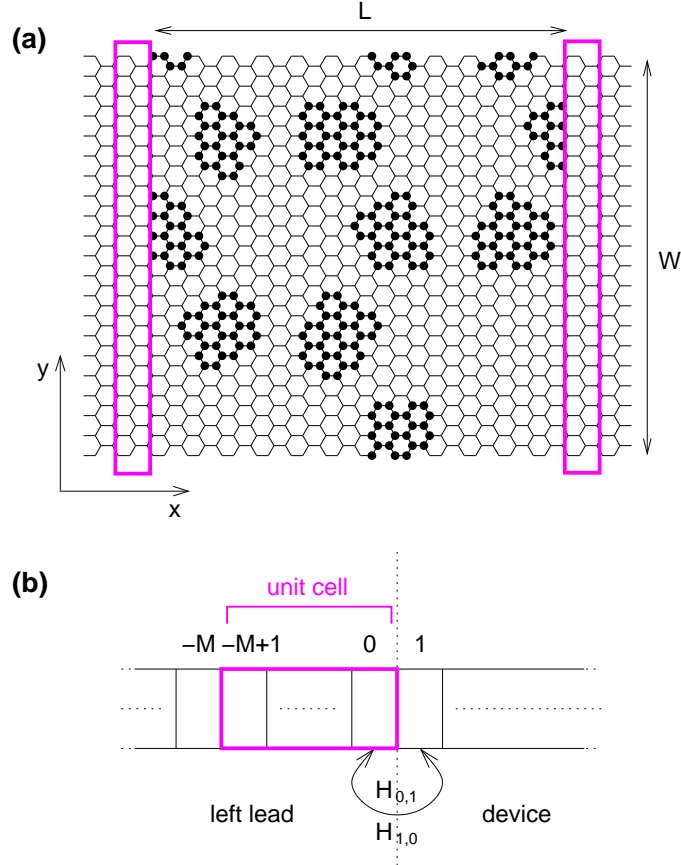


FIG. 14: (a) Example of device used in transport simulations: hydrogenated channel of size $W \times L$ between two leads of pristine graphene. The device is aligned so that its longitudinal direction corresponds to an armchair direction. (b) Slab representation of the same structure. The unit cell inside the lead regions can be viewed as being made of $M = 4$ slabs, where each slab corresponds to a row of carbon atoms.

with

$$\Omega^C = \begin{pmatrix} \Omega_{-M+1,-M+1} & \Omega_{-M+1,-M+2} & & \\ \Omega_{-M+2,-M+1} & \ddots & \ddots & \\ & \ddots & \Omega_{-1,-1} & \Omega_{-1,0} \\ & & \Omega_{0,-1} & \Omega_{0,0} \end{pmatrix}. \quad (\text{F6})$$

As a first step, we consider Ω^C and decimate all the slabs from -1 backward to $-M+2$ (assuming $M > 2$). We define $d_1^{(0)} = \Omega_{0,0}$, $d_2^{(0)} = \Omega_{-1,-1}$, $a^{(0)} = \Omega_{-1,0}$, $b^{(0)} = \Omega_{0,-1}$. The generic iteration of index n ($n = 1, \dots, M-2$) consists in eliminating the second last node from the matrix

$$\begin{pmatrix} \ddots & & \ddots & & \\ \ddots & \Omega_{-n-1,-n-1} & \Omega_{-n-1,-n} & & \\ & \Omega_{-n,-n-1} & d_2^{(n-1)} & a^{(n-1)} & \\ & & b^{(n-1)} & d_1^{(n-1)} & \end{pmatrix} \quad (\text{F7})$$

with the equations

$$\begin{aligned} d_1^{(n)} &= d_1^{(n-1)} - b^{(n-1)} \left(d_2^{(n-1)} \right)^{-1} a^{(n-1)}, \\ d_2^{(n)} &= \Omega_{-n-1,-n-1} - \Omega_{-n-1,-n} \left(d_2^{(n-1)} \right)^{-1} \Omega_{-n,-n-1}, \\ a^{(n)} &= -\Omega_{-n-1,-n} \left(d_2^{(n-1)} \right)^{-1} a^{(n-1)}, \\ b^{(n)} &= -b^{(n-1)} \left(d_2^{(n-1)} \right)^{-1} \Omega_{-n,-n-1}, \end{aligned} \quad (\text{F8})$$

which are simply an application of (C5). At the end, we obtain the renormalized Ω^L matrix

$$\tilde{\Omega}^L = \begin{pmatrix} \ddots & & \ddots & & \\ \ddots & d_2^{(M-2)} & a^{(M-2)} & & \\ & b^{(M-2)} & d_1^{(M-2)} & \Omega_{-M,-M+1} & \\ & & \Omega_{-M+1,-M} & d_2^{(M-2)} & a^{(M-2)} \\ & & & b^{(M-2)} & d_1^{(M-2)} \end{pmatrix}. \quad (\text{F9})$$

As a second step, we consider $\tilde{\Omega}^L$ and decimate all the even slabs (assuming $M > 1$). By using the formulas (again an application of Eqs. C5)

$$\begin{aligned} \delta_1^{(0)} &= d_1^{(M-2)} - b^{(M-2)} \left(d_2^{(M-2)} \right)^{-1} a^{(M-2)}, \\ \delta_2^{(0)} &= \delta_1^{(0)} - \Omega_{-M,-M+1} \left(d_2^{(M-2)} \right)^{-1} \Omega_{-M+1,-M}, \\ \alpha^{(0)} &= -\Omega_{-M,-M+1} \left(d_2^{(M-2)} \right)^{-1} a^{(M-2)}, \\ \beta^{(0)} &= -b^{(M-2)} \left(d_2^{(M-2)} \right)^{-1} \Omega_{-M+1,-M}, \end{aligned} \quad (\text{F10})$$

we get a new renormalized Ω^L matrix,

$$\tilde{\tilde{\Omega}}^L = \begin{pmatrix} \ddots & & \ddots & & \\ \ddots & \delta_2^{(0)} & \alpha^{(0)} & & \\ & \beta^{(0)} & \delta_2^{(0)} & \alpha^{(0)} & \\ & & \beta^{(0)} & \delta_2^{(0)} & \alpha^{(0)} \\ & & & \beta^{(0)} & \delta_1^{(0)} \end{pmatrix}. \quad (\text{F11})$$

This matrix has the same structure as the one used in the Sancho-Rubio algorithm [25]. The generic iteration of index n ($n = 1, 2, \dots$) of this algorithm actually consists in the decimation of the slabs with even indexes from the matrix

$$\begin{pmatrix} \ddots & & \ddots & & \\ & \ddots & \delta_2^{(n-1)} & \alpha^{(n-1)} & \\ & & \beta^{(n-1)} & \delta_2^{(n-1)} & \alpha^{(n-1)} \\ & & & \beta^{(n-1)} & \delta_2^{(n-1)} & \alpha^{(n-1)} \\ & & & & \beta^{(n-1)} & \delta_1^{(n-1)} \end{pmatrix}, \quad (\text{F12})$$

by using the formulas (again from Eqs. C5)

$$\begin{aligned} \delta_1^{(n)} &= \delta_1^{(n-1)} - \beta^{(n-1)} \left(\delta_2^{(n-1)} \right)^{-1} \alpha^{(n-1)}, \\ \delta_2^{(n)} &= \delta_2^{(n-1)} - \beta^{(n-1)} \left(\delta_2^{(n-1)} \right)^{-1} \alpha^{(n-1)} - \alpha^{(n-1)} \left(\delta_2^{(n-1)} \right)^{-1} \beta^{(n-1)}, \\ \alpha^{(n)} &= -\alpha^{(n-1)} \left(\delta_2^{(n-1)} \right)^{-1} \alpha^{(n-1)}, \\ \beta^{(n)} &= -\beta^{(n-1)} \left(\delta_2^{(n-1)} \right)^{-1} \beta^{(n-1)}, \end{aligned} \quad (\text{F13})$$

until convergence, i.e. until the coupling matrices $\alpha^{(n)}$ and $\beta^{(n)}$ become sufficiently small. At the end, we can approximate $g_{0,0}^r = \left(\delta_1^{(n)} \right)^{-1}$, where n stands for the index of the last iteration.

-
- [1] A. K. Geim and K. S. Novoselov, *Nature Materials* **6**, 183 (2007).
 - [2] A. H. C. Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, *Rev. Mod. Phys.* **81**, 109 (2009).
 - [3] M. Y. Han, B. Özyilmaz, Y. Zhang, and P. Kim, *Phys. Rev. Lett.* **98**, 206805 (2007).
 - [4] X. Wang, Y. Ouyang, X. Li, H. Wang, J. Guo, and H. Dai, *Phys. Rev. Lett.* **100**, 206803 (2008).
 - [5] T. G. Pedersen, C. Flindt, J. Pedersen, N. A. Mortensen, A.-P. Jauho, and K. Pedersen, *Phys. Rev. Lett.* **100**, 136804 (2008).
 - [6] J. Bai, X. Zhong, S. Jiang, Y. Huang, and X. Duan, *Nature Nanotechnology* **5**, 190 (2010).
 - [7] R. Balog, B. Jørgensen, L. Nilsson, M. Andersen, E. Rienks, M. Bianchi, M. Fanetti, E. Lægsgaard, A. Baraldi, S. Lizzit, et al., *Nature Materials* **9**, 315 (2010).
 - [8] J. O. Sofo, A. S. Chaudhari, and G. D. Barber, *Phys. Rev. B* **75**, 153401 (2007).
 - [9] D. C. Elias, R. R. Nair, T. M. G. Mohiuddin, S. V. Morozov, P. Blake, M. P. Halsall, A. C. Ferrari, D. W. Boukhvalov, M. I. Katsnelson, A. K. Geim, et al., *Science* **323**, 610 (2009).
 - [10] G. Fiori, S. Lebegue, A. Betti, P. Michetti, M. Klintonberg, O. Eriksson, and G. Iannaccone, *Phys. Rev. B* **82**, 153404 (2010).
 - [11] R. Petersen, T. G. Pedersen, and A.-P. Jauho, *ACS Nano* **5**, 523 (2011).
 - [12] M. Yang, A. Nurbawono, C. Zhang, Y. P. Feng, and Ariando, *Appl. Phys. Lett.* **96**, 193115 (2010).
 - [13] J. P. Robinson, H. Schomerus, L. Oroszlány, and V. I. Fal'ko, *Phys. Rev. Lett.* **101**, 196803 (2008).
 - [14] J. Bang and K. J. Chang, *Phys. Rev. B* **81**, 193412 (2010).
 - [15] A. T. N'Diaye, S. Bleikamp, P. J. Feibelman, and T. Michely, *Phys. Rev. Lett.* **97**, 215501 (2006).
 - [16] R. Brako, D. Šokčević, P. Lazić, and N. Atodiresei, *New J. Phys.* **12**, 113016 (2010).
 - [17] V. M. Pereira, J. M. B. L. dos Santos, and A. H. C. Neto, *Phys. Rev. B* **77**, 115109 (2008).
 - [18] D. W. Boukhvalov, M. I. Katsnelson, and A. I. Lichtenstein, *Phys. Rev. B* **77**, 035427 (2008).
 - [19] Y.-W. Son, M. L. Cohen, and S. G. Louie, *Phys. Rev. Lett.* **97**, 216803 (2006).
 - [20] D. Gunlycke, D. A. Areshkin, and C. T. White, *Appl. Phys. Lett.* **90**, 142104 (2007).
 - [21] M. Y. Han, J. C. Brant, and P. Kim, *Phys. Rev. Lett.* **104**, 056801 (2010).
 - [22] P. W. Anderson, D. J. Thouless, E. Abrahams, and D. S. Fisher, *Phys. Rev. B* **22**, 3519 (1980).
 - [23] X. Li, W. Cai, J. An, S. Kim, J. Nah, D. Yang, R. Piner, A. Velamakanni, I. Jung, E. Tutuc, et al., *Science* **324**, 1312 (2009).
 - [24] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, 1997).
 - [25] M. P. L. Sancho, J. M. L. Sancho, and J. Rubio, *J. Phys. F: Met. Phys.* **15**, 851 (1985).
 - [26] T. Low, S. Hong, J. Appenzeller, S. Datta, and M. S. Lundstrom, *IEEE Trans. Elec. Dev.* **56**, 1292 (2009).
 - [27] P. A. Lee and T. V. Ramakrishnan, *Rev. Mod. Phys.* **57**, 287 (1985).
 - [28] M. Jaiswal, C. H. Y. X. Lim, Q. Bao, C. T. Toh, K. P. Loh, and B. Özyilmaz, *ACS Nano* **5**, 888 (2011).

- [29] A. L. V. de Parga, F. Calleja, B. Borca, J. M. C. G. Passeggi, J. J. Hinarejos, F. Guinea, and R. Miranda, Phys. Rev. Lett. **100**, 056807 (2008).
- [30] G. Grosso, S. Moroni, and G. P. Parravicini, Phys. Rev. B **40**, 12328 (1989).
- [31] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, J. Appl. Phys. **81**, 7845 (1997).
- [32] T. Low and J. Appenzeller, Phys. Rev. B **80**, 155406 (2009).
- [33] P. Zhao and J. Guo, J. Appl. Phys. **105**, 034503 (2009).
- [34] As explained in Ref. 13, the effect of each hydrogen atom at energy E can be recast in a effective self-energy $\Sigma_H = \gamma_H^2 / (E + i0^+ - \epsilon_H)$ for the attached carbon orbital: since $\Sigma_H \gg 1$ eV for $0 < |E| < 1$ eV, the hydrogenated carbon atom effectively acts as a vacancy in the energy range of interest.
- [35] It can be divided into two regions that are equivalent to each other under reflection, apart from the exchange of the two graphene sublattices.
- [36] More precisely, within the TB representation, the continuous position \mathbf{r} is substituted by the pair (\mathbf{l}, q) , with \mathbf{l} the graphene lattice vector and q the orbital index: the projected spectral function for carbon atoms only is considered ($q = 1, 2$), the Fourier transform is done with respect to \mathbf{l} , and the result is summed over q , that is over the two graphene sublattices (an additional factor is used in our plots to normalize the number of states with respect to the graphene unit cell; spin degeneracy is not included).
- [37] We note the absence of repeating dispersions expected for a periodic superlattice structure. This is due to the effect of disorder, which destroys the strict periodicity.
- [38] For constant supercell size and filling factor, the points seem to depart from the linear trend. This could be explained by the fact that, for some ensembles, the amount of disorder on the cluster edges is not strong enough to completely wash out the interference effect of the specific cluster shape. This situation is particularly evident for $N_w = 3$, where the repeated band structure of the superlattice is still distinguishable in the $A(\mathbf{k}, \mathbf{k}; E)$ plot (see Fig. 2a-right and additional plots in Supp. info).
- [39] We note that, since the transport simulations confirm the previously extracted gap size, the gap is expected to have the same extension in other directions than the one used for Figs. 1c, 2a.
- [40] <http://www.caam.rice.edu/software/ARPACK/>